

For the Europe of 23 languages: the computer learns to translate

The new European project EuroMatrix develops automatic translation systems for all official languages of the EU.

The number and scope of translations in business, administration, politics and culture is increasing yearly. The principle clients for translations are export-oriented sectors in industry, international organisations including the European Union and the United Nations, as well as multinational corporations. The exploding demand for text translations is one of the direct consequences of open trade practices, globalisation, European integration and international cooperation. The legion of translators keeps growing. However, with the general trend towards specialisation, the demand on the competence of expert translators also increases. This leads to a general increase in the cost of a high quality translation. The European Commission cannot, for example, send technical and legal translations between the languages of the European Union to the Philippines, India or other low-wage countries.

A solution is expected from research on automatic translation, which is often presented as being the only realistic solution to managing the communication problems of a multilingual world society. Automatic translation is increasingly viewed by experts as a critical precondition for the sustainable preservation of the linguistic and cultural diversity on our planet.

Although experts agree among themselves on the immense demand for automatic translation, the opinions concerning the possibilities of such systems with regard to complex texts differ widely. While many experts still doubt that machines can deliver adequate translations, there are at the same time corporations already reducing costs through the smart utilization of existing translation technology.

The truth of the matter lies not quite in the middle, although both sides are right in their own way. Even outdated and low-quality translation systems, which are currently only of very limited use in business and administration, bring costs down. The real breakthrough in automatic translation, however, still lies before us. After a long period of stagnation, a reason for optimism is to be found in the impressive research being done in statistical translation. These programs learn translation patterns from large bodies of already-translated text. They are still far from being perfect, but they are improving from year to year.

However, they make different errors to those made by the conventional, rule-based translation programs, designed by linguistics. Thus a solution may be found in a clever combination of the two methods.

Technological breakthroughs in automatic translation are to be the aim of a new international research program which is being financed under the 6th Framework Program of the European Union. New combinations of the best currently available methods for rule-based and statistical translation are to be tested. Additionally, the question of how statistical translation systems can be improved by linguistic methods will be investigated. The ambitious aim is to develop and test translation systems for the 23 official languages of the European Union. An important part of the project is also the organization of competitions among the best translation systems, to which all research centers and firms from Europe and other parts of the world are to be invited to participate. It will naturally not be possible to focus equally on each language pair, but a practical result of the project is to be a continuously updated appraisal of

the technological state for each language. This overview in the form of a large tabular matrix has given the project its name.

The project is coordinated by Saarland University represented by Professor Hans Uszkoreit. Dr. Philipp Koehn from the University of Edinburgh is serving as technical/scientific coordinator of the project. Other partners include internationally renowned research groups at Charles University in Prague lead by Professor Jan Hajic and the Center CELCT in Trento under the direction of Professor Amedeo Cappelli.

Important for the success of the project is the participation of industrial partners: the German firm GroupWare AG is bringing their LOGOS system into the project, LOGOS being one of the largest and most important commercial translation systems, which was made available the year before last in an Open-Source version. The firm Morphologic in Budapest has developed digital dictionaries for many European languages and has its own translation technology.

The project has recently kicked off and is planned for 30 months at a cost of 2.5 million Euros.

Contact: Dr. Andreas Eisele
 Saarland University
 eisele@dfki.de